# Multiresolution Information Archival and Analysis System

**Donald Fussell**

**Al Bovik**

**Melba Crawford**

**Harrick Vin**

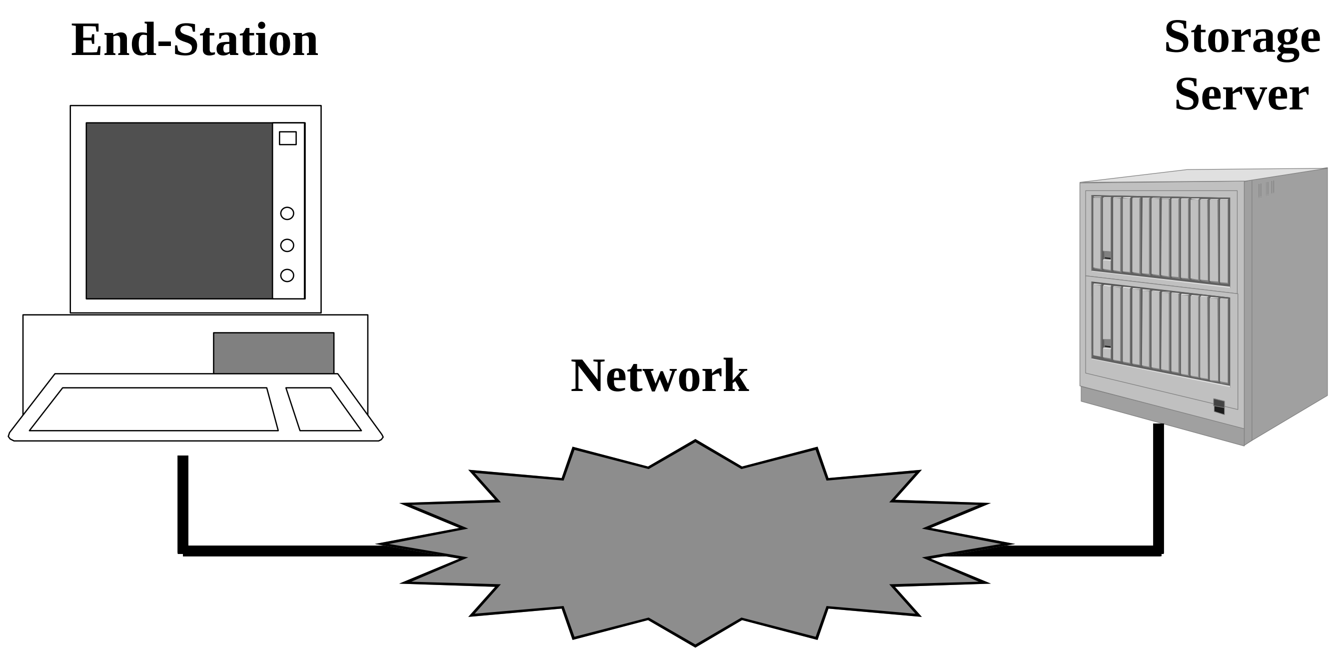
**The University of Texas at Austin**

**NASA EOSDIS Technology Transfer Workshop - February 1996**

# Managing the Quality of Query Results

- **The need for variable quality in EOSDIS**
  - **The nature of the data: very large number of voluminous objects**
  - **The consequent high cost of storage, transmission, and processing**
  - **Diversity in users' quality requirements**
  - **Diversity in users' storage, transmission and processing capabilities**
- **Multiresolution as a means of managing quality**

# Objective




**Design and implement an end-to-end architecture for distributed EOSDIS applications**

# Key Challenge: Managing Heterogeneity

- **Computing: ranging from high-end workstations to hand-held devices**

- **Communication: ranging from high-bandwidth ATM networks to low-bandwidth wireless environments**

- **Data type and formats (e.g., various compression algorithms)**

# Research Horizon

- Design and implement a multiresolution multimedia file system
- Develop algorithms and protocols for efficient transmission of multimedia data over networks
- Design and implement OS and transport protocol related services at the end-station
- Design and implement a tool for accessing multiresolution EOSDIS data by by various user communities

# Multi-resolution Multimedia File System

- Supports efficient storage and retrieval of multiple resolution levels of each object (containing images, video, etc.)
- Supports periodic and aperiodic real-time, as well as non real-time requests
- Understands and provides various levels of Quality of Service (QoS) to clients

# Key Features

- **Integrated data management**
  - **Data type specific:**
    - **Storage structures (e.g., i-nodes)**
    - **Placement and retrieval algorithms**
    - **Failure recovery mechanisms**
- **Scheduling algorithms:**
  - **Minimize response times for non real-time requests**
  - **Meet the performance requirements of real-time requests**
- **Adapts QoS to changes in resource availability**

- **Design principle: achieve efficient system design by exploiting the semantics of data**

# Failure Recovery in Disk Arrays

- **Disk mirroring**
  - **100% storage space overhead**
- **Parity encoding: Recovery using redundant parity information**
  - **balanced system load prior to failure => 100% increase in load on surviving disks after failure**
  - **Prevent saturation => low utilization in fault-free state**
  - **Declustered parity reduces overhead to (G-1)/(C-1) where G : parity group size and C: cluster size**

| disk 1 | | | disk 4 | |
|---|---|---|---|---|
| | | M0.2 | M0.3 | |
| | | M1.2 | P1 | |
| | | P2 | M2.2 | |
| | | M3.1 | M3.2 | |
| | | M4.1 | M4.2 | |

# Failure Recovery in Multimedia Servers

- **Observations:**
  - Images are inherently redundant
  - Human perception is tolerant to minor distortions in visual presentations
- **Approach:**
  - Improve the performance by exploiting properties of media streams

# Addressing Scalability

- **Clustered file system architecture**
  - **Cluster: collection of nodes connected by a network**

- **Developed techniques for:**
  - **Replicating and/or distributing the storage of each object across multiple nodes**
  - **Tolerating disk and node failures**

# Implementation Status

- Basic multi-user file system (in user-space) is operational
- Resource reservation and management, as well as failure recovery methods are being implemented
- Experiments with clustered file system are being carried out on an IBM SP-2
- Plan:
  – Complete user-sapce file system implementation by June 1996
  – Migrate it to kernel by December 1996

# Network Algorithms and Protocols

- Efficient transmission of compressed imagery over networks
- Key objective: combine multiresolution encoding with prioritized and progressive transmission schemes
- Supports:
  - Recovery from error in transmission (e.g., packet loss) without retransmission
  - Negotiation for appropriate quality of service
  - On-the-fly changes in quality using QoS filters

# Status and On-going Work

- **Status:**
  - **Implemented and experimented with efficient techniques for transmitting images compressed usng a DCT-based algorithm**
- **Ongoing work:**
  - **Extensions for wavelet- and fractal-based techniques**
  - **Algorithms and protocols for integrated traffic management**
  - **Framework for hierarchical resource management**

# OS and Transport Services

- **Quality of Service (QoS) architecture**
  - **Efficient mechanisms for transport and presentation-level processing**
  - **Framework for meeting the QoS requirements of applications**

- **Framework should support:**
  - **Specification of QoS requirements**
  - **On-demand composition to adapt to:**
    - **Resource availability**
    - **Application requirements**

# Example: Presentation Processing Engine (PPE)

- **Library of configurable compression and image processing modules**
- **Modules can be dynamically configured to create a presentation processing engine**
- **Modular architecture**
  - **Software engineering benefits**
  - **Possible to incorporate image processing functions to manipulate compressed data => substantial improvement in performance**

# Distributed Query Interface

- Dynamic user-specified QoS requirements
  - Cost/performance tradeoffs specified as value functions - part of query interface
- Online feedback from storage and transport components for dynamic resolution adjustment
  - Provide best possible quality of data meeting performance constraints
  - Provide required data with best available performance given priority
- Browse dynamically enhanced data

# Current Status

- Implementation based on "sandbag" multiresolution data model
- "sandtree" data structure supports aggregate operations and multiresolution image queries efficiently
- Client-server architecture
- Simple value functions - direct resolution specification or time bound specification
- Crude system performance estimates

# Aspects of "Quality"

- **Information detail contained in query results**
  - **Precision of data values: direct, correlated**
  - **Density of data points: spatial, temporal, band**
  - **Membership completeness of data sets**
  - **Abstraction level of data**
- **The strength of guarantees on query results**
  - **Currency**
  - **Consistency**

# Methods for EOSDIS

- Issues in choosing a method of generating multiple resolution versions
  - Ease of generation of multiple resolutions
  - Effectiveness of cost reduction: storage, transport, data management
  - Processing algorithm “friendliness” of various resolution versions
- Managing multiple resolution versions
  - Storage of precomputed versions, on-demand computation, customized generation based on user-supplied scripts

# Using Multiresolution Versions

- **Constructing cost models based on multiresolution**
  - **Overall cost a function of resolution along each aspect of quality**
- **Factors in selecting the resolution of result for a given query**
  - **Direct specification in query, bound on cost, bounds on storage size and delivery time, server load**

# A Model for Distributed Servers

- Autonomous administration of individual servers
- Information disseminated via query interfaces in servers
- Possible server interactions: propagation and exchange of information
- Absence of global transactions

# A Model of Augmented Information Objects

- No global correctness, currency, or consistency management protocol
- Information objects augmented with meta-data
- Basic assumption: information object and associated meta-data were correct together at some point in time
- Correctness, currency and consistency properties of a query result derived from meta-data of objects in result set

# Definitions of Correctness, Currency, and Consistency

- Correctness with respect to
  - the object being modeled
  - the totality of information servers
  - the answering server
- Currency definitions based on visibility of changes/updates
- Consistency definitions based on duration of simultaneous correctness

# Evaluation of the Autonomous Server Model

- Advantages over a traditional DBMS
  - More autonomy for individual servers
  - Reduced protocol overhead in object exchange
  - Higher tolerance of unavailability of servers
- Disadvantages
  - Weaker guarantees on query results

# Basic Meta-data Scheme for EOSDIS Servers

- **Meta-data for objects:**
  - **Spatial, temporal and grid information**
  - **Content description and classification**
  - **Processing history and available resolutions**
- **Meta-data for servers:**
  - **Aggregate object descriptions**
  - **Meta-data on server capacity and data management parameters**

# Issues in Meta-data Management for EOSDIS

- **Interoperability: Adopting a general format for meta-data**
- **Multiresolution in meta-data and its effects on search and navigation efficiency**
- **Partially automated derivation of meta-data hierarchy from object meta-data**

# Multiresolution Image Analysis

- **Goal - Develop and apply methods for image analysis which both exploit and test multiresolution datga structures, coding, storage, and transport protocols.**
- **Focus - Classes of approaches which will be required to satisfy operational requirements of EOSDIS data (i.e. retrieval of large data sets and operational data products by multiple users over networks.**
- **Characteristics of problems included in UT study:**
  - **Enhancement of multispectral/multitemporal data for visual browsing and analysis**
  - **Analysis of multiple data sets with different spatial/spectral characteristics**
  - **Analysis of noisy data requiring identification and possible removal of anomalies**
  - **Extraction and possible tracking of features**
  - **Segmentation and classification of imagery**

- **Specific Project Applications at UT CSR being utilized for system evaluation**
  - **Image registration (multispectral and SAR data from SPOT, Landsat, ERS-1, SIR-C, Airborne CAMS, JPL AIRSAR)**
  - **Interpolation of altimetry data for geoid mapping and tracking dynamic topography (GEOSAT, ERS, and TOPEX)**
  - **Reconstruction of hundreds of daily AVHRR data for vegetation mapping and change detection studies (UT receiving station)**
  - **Integration of AVHRR and GOES imagery for regional mesoscale meteorological studies (UT HRPT and Stgate of TExas GOES Seaspace/Lockheed systems)**
  - **Automatic extraction of feature boundaries from full Landsat scenes for fire mapping (25 years of MSS data over W. Australia)**
  - **Tracking thermal ocean features in AVHRR imagery (UT HRPT station)**
  - **Landscape simulation for habitat studies**
  - **Topographic studies (interferometry and stereo combinations from ERS, SIR-C and SPOT imagery)**

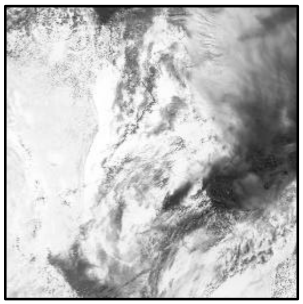
# Multiresolution Analysis Approches

- **Multitemporal Image Reconstruction**
  - **Markov random field spatial model coupled with a time series model implemented within a pyramid structure to reconstruct sequences of hundreds of images (each on the order of 2kx2k)**
- **Multiresolution Registration**
  - **Wavelet-based decomposition combined with feature extraction and matching**
- **Multiresolution Spatial Interpolation**
  - **Image pyramid combined with adaptive Kalman filter**
- **Multiresolution Simulation**
  - **Markov random field model used to simulate large region process, then boundary variation and intraregion characteristics superimposed via pyramid structure**
- **Feature Extraction and Tracking**

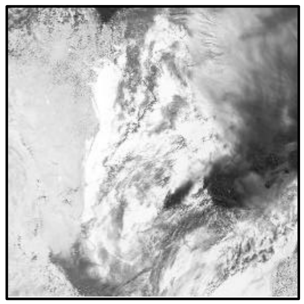
- **Anisotropic diffusion pyramid developed to detect boundaries and track features at multiple resolutions**

- **Multiresolution Segmentation and Classification**
  - **Wavelet-based decomposition combined with multiresolution Markov random field model**

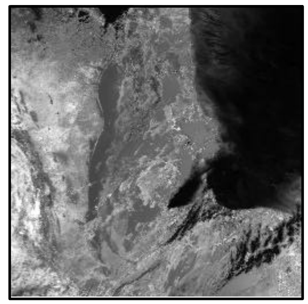
# Multispectral Image Data



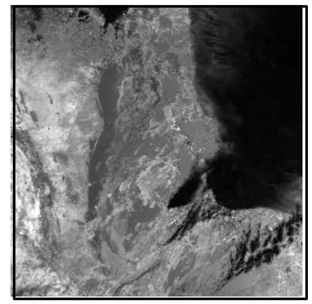
Band 1



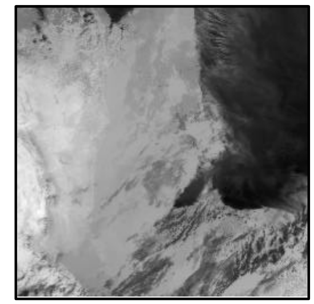
Band 2



Band 3



Band 4



Band 5

- **Images may have high spatial resolution, wide dynamic range, and hundreds of bands**
- **Data sets are extrmely large, requiring expensive storage and long network transmission times**
- **Lossless and lossy compression schemes needed**

# Compression Schemes

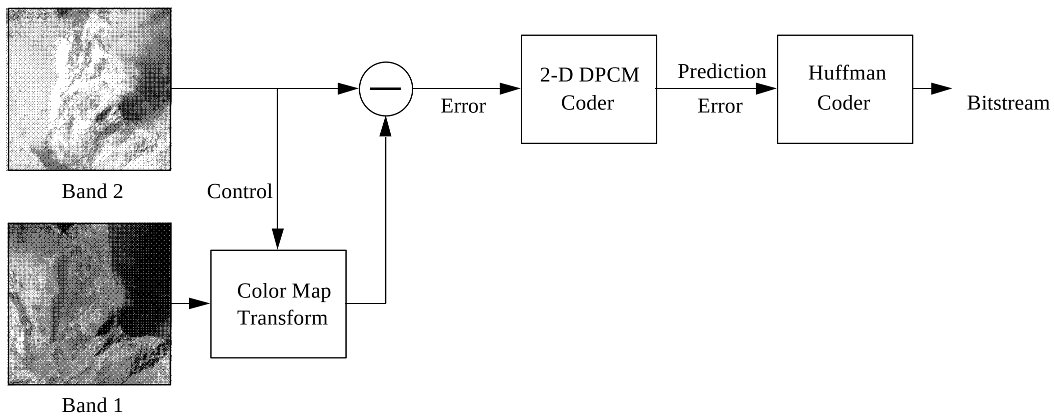
## LOSSLESS

- **Required for archiving and machine vision**
- **Relies on reduction of entropy of the image**
- **Achievable compression is low (2:1 typical for single band images)**

## LOSSY

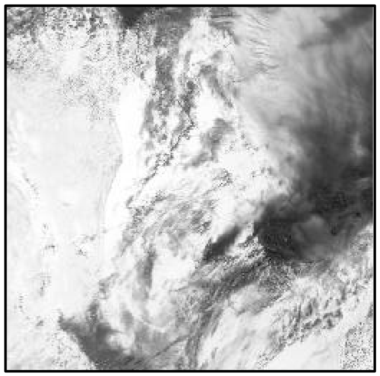
- **Relies on properties of human visual system**
- **High quality images at high compression ratios are obtainable**
- **Good for transmission over networks, non-critical applications**

# Lossless Image Coding

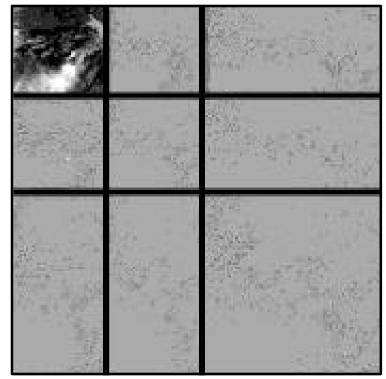



- **Typical data has high spectral/spatial redundancy**
- **Spectral correlation is reduced using the CMT**
- **Spatial correlation is reduced using DPCM**
- **Variable-length coding approaches entropy bound**

# Lossy Wavelet Image Coding



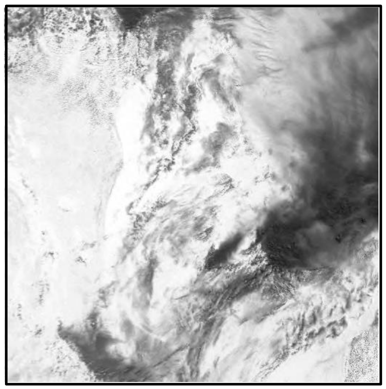
Original data



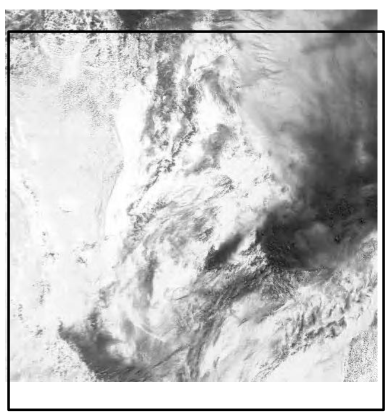
One possible wavelet decomposition

- **Image is split into parallel frequency channels**
- **Natural images are inherently multiresolution**
- **Human visual system uses frequency channels**
- **Efficient storage and progressive transmission**

# Single Band Compression



EPIC coder 37:1



Wavelet coder 37:1

- **Efficient wavelet coders exist but for single bands only**
- **Good visual quality at 30:1 available**
- **Multispectral data has higher redundancy and will allow higher compresson ratios**